

DOCUMENT RESUME

ED 041 359

EA 002 905

AUTHOR Feldman, Kenneth A.
TITLE Using the Work of Others: Some Observations on
Reviewing and Integrating.
PUB DATE [69]
NOTE 23p.
EDRS PRICE MF-\$0.25 HC-\$1.25
DESCRIPTORS Educational Research, *Educational Sociology,
*Literature Reviews, *Research Methodology, Research
Problems, *Research Reviews (Publications), *Social
Sciences

ABSTRACT

Systematically reviewing and integrating the literature of a field of study may be considered a type of research using a particular set of research techniques and methods. The works of other writers comprise the integrator's raw data. Using bibliographies and reviews, he must locate all references that seem relevant to a given topic. Special problems include determining what material to integrate, devising ways of comparing diverse studies, resolving difficulties arising as a result of missing data, and reporting analyses well. An integration of literature and studies in a particular field serves as a compendium of research, compiling previously dispersed information, abstracting existing research, and providing bibliographic listings of research and theory. Such an integration also simulates replication by comparing similar studies. Additionally, it generates propositions, provides for extended statistical analysis of completed studies, and clarifies conflicts between schools of thought. The author's comments are supported by references from the field of educational sociology. (JK)

ED041359

N-X

**USING THE WORK OF OTHERS: SOME OBSERVATIONS
ON REVIEWING AND INTEGRATING**

**Kenneth A. Feldman
State University of New York
at Stony Brook**

**U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE**

OFFICE OF EDUCATION

**THIS DOCUMENT HAS BEEN REPRODUCED
EXACTLY AS RECEIVED FROM THE PERSON OR
ORGANIZATION ORIGINATING IT. POINTS OF
VIEW OR OPINIONS STATED DO NOT NECES-
SARILY REPRESENT OFFICIAL OFFICE OF EDU-
CATION POSITION OR POLICY.**

EA 002 905

There are many ways of social sciencing. Large-scale review and integration of existing theory and research in some bounded area is one such way. Systematically reviewing and integrating what is nominally called the "literature" of a field may be considered a type of research in its own right--one using a characteristic set of research techniques and methods. This paper highlights some of the processes, problems, functions and values of this type of research, with special reference to the domain of educational sociology.¹

PROCEDURES AND PROBLEMS OF INTEGRATIONS

The written work of others (whether or not formally published) constitutes "raw" data for the integrator. In compiling the initial list of materials to be read, the integrator makes use of other bibliographies, reviews and integrations in the specific areas of his concern. In addition, there are a variety of more general sources of information.² If appropriate resources are available, it is also advisable for the integrator to write to individuals and institutes who are doing (or have done) research in the area being integrated--asking for materials on the subject, and requesting any bibliographic aid they might be able to supply (including the titles of reports and the names of researchers that might otherwise be passed by). Further, as any given report is read, the conscientious integrator will make it a point to track down any references given in it that seem relevant to his topic. After a given period--it may be a long one--little in the way of "new" references will be found.

This general collection procedure entails a sampling among theoretical and empirical efforts--in two different ways. The first kind of sampling is substantive. From all possible topics of concern, the integrator delimits the general area of theory and research to be considered, and spells out relevant sub-topics and phenomena. He will probably want to keep the range a little wider, and his definitions of relevant concerns more general, than is absolutely necessary. In so doing he will find himself skimming literature he will not fully need for his final analysis, but such a procedure will help him not to overlook important studies, and will furnish him with a broader background for his analysis.

Having established the substantive boundaries and specified the topics of his integration, the materials that the integrator then collects can only be a sample of existing studies. Although his collection procedures will lead to many unpublished reports and other materials not generally available or easily obtainable (for example, doctoral dissertations), they do not by any means guarantee total coverage. To attempt total collection of all possible materials--including undergraduate research reports written for a class; master's thesis; unpublished "in-house" reports of teachers, deans, principles, counselors, and directors of research institutes; unpublished memos, and the like--is foolhardy (and impossible, to boot). The integrator should not intend a "flight into exhaustiveness" (just as he avoids an "exhaust into flightiness"). The integrator will limit himself to covering research done within a given time period, and perhaps also to certain kinds of written materials. Within these limits, every integrator must hope to be reasonably, and therefore satisfactorily, comprehensive.

The difficulty involved in this sampling of studies, of course, is that the parameters of the universe are not known, and thus the representativeness of the sample cannot be determined. The integrator hopes that by obtaining his bibliographic references from a wide variety of sources, he can discover less publicized research and avoid a too-biased selection of studies (say, only those by investigators in the "main stream," or only those subject to current fads, distortions, etc.) He may also assume that the larger his sample (the more studies he has tracked down), the more representative and trustworthy it is, but this can only be a hope on his part. Thus the conclusions of his research must always be somewhat tentative.

Certainty must be balanced with uncertainty. For example, with only slight exceptions, nearly three dozen studies have found that college seniors when compared to freshmen (either cross-sectionally or longitudinally) are typically less authoritarian, dogmatic, prejudiced and stereotyped in their thinking (Feldman and Newcomb, 1969). Results are not only usually statistically significant with a study, but consistent across studies. Such a set of findings leads to a type of certainty with respect to conclusions in this area. It is always possible,

however, that there exists a number of reports (perhaps little known or even unknown), that do not show freshman-senior decreases on these variables or actually show freshman-senior increases. Although this is highly unlikely, possibilities such as these must be kept in mind. The integrator, then, makes the best generalizations he can with the studies he has managed to collect--always realizing that his conclusions are open to challenge from past studies he has missed (not to mention future studies he has no way of getting).

The amount of information available in the typically tens or hundreds of research reports in an area being integrated would be overwhelming unless the integrator develops a systematic schema of indexing, coding, and retrieving this information. One way of doing this is to generate a set of topics or categories of particular relevance, with which to index and code the materials he is reading. For some integrations it may be possible to completely extract, categorize, and code the information in a book or article the first time it is read; in this case it is not necessary to return to the original source. In other instances, the integrator will find it more fruitful to initially read the materials he is collecting, make a note of the existence of the part(s) of the report that deals with topics relevant to his integration, and perhaps write a short comment about the researcher's findings. When it comes time to integrate the materials on a particular topic he can then gather together his notes, his initial analyses, and the articles and reports themselves, for re-reading, for subdividing into sub-topics, and for further analyses.

One of the major tasks for the integrator is to devise means of meaningfully comparing diverse studies. Sometimes so-called reviews of the literature do no more, essentially, than string together short summaries of selected studies. This procedure does have its worth, but it only minimally qualifies as a "review," and certainly in no way can be considered an "integration" of the literature. I am reminded of the Vaughan Meader comic recording of a few years back in which he ostensibly asked the then first lady if she would be kind enough to point out the paintings in the White House. She said she would be glad to, and proceeded through the White House, commenting at appropriate locations, "Here's a painting, and here's

a painting, and here's a painting..." Pointing out studies in this manner--even when accompanied by short descriptions--leaves the observer with the same feeling of emptiness.

Beading together a series of abstracts, while of some value, is an inadequate device to compare studies. Other techniques of comparison can be developed. One simple procedure is to explicitly count the direction of results of researches in a given subarea: of these twelve studies, eight support a certain theory or hypothesis and the other four do not. Hopefully, the integrator is then able to use the studies to go past mere counting, and say something about the conditions under which certain generalizations hold and the conditions under which they do not.³ As illustrated below, other means of comparing studies can also be devised.

I should like to discuss in greater detail certain problems and processes involved in comparing studies by describing one part of an integration (commissioned and financially supported by the Carnegie Foundation for the Advancement of Teaching) that Theodore M. Newcomb and I attempted on the impacts of colleges on students (Feldman and Newcomb, 1969). In our search of the literature, we found a large number of investigations that in some way measured, described, and compared students in different major fields. We wanted to compare the results of these studies in a systematic way.

For meaningful comparisons across studies, the attributes on which students were compared were divided into clusters, which permitted considering together only those studies dealing with a given cluster. Such division is arbitrary, but does not need to be meaningless. We were guided by previous theoretical and empirical work, and by the scales and instruments actually used in the studies themselves.

Among the studies researching curricular differences, classification of major fields varies. Some studies use a classification scheme that is at the departmental level. Others classify curricula at the divisional school level. Some mix the two kinds. Even those using the same general level of classification, say departmental, do not use exactly the same categories or combination of categories. This meant that we were forced to use categories of comparison that combined departments in

a logical way given the existing classifications in the studies: for example, "English, or Literature, or English and Languages, or Literature and Languages."

Most of the studies in this area present average scores of students. But some show the percentage of students in each curricular group who score a certain way on a scale (for example, the percentage who get a "high" score); and others rank the fields in some way. Also, studies vary in the way in which they test these differences for statistical significance. Some, through pair-by-pair comparisons, test the significance of the difference between any one curriculum and each of the other curricula. Others, by using statistical techniques such as analysis of variance, merely determine whether the means of major fields differ significantly among themselves, but do not indicate which of the specific fields are significantly different from each other. Many studies do not test for statistical significance of differences at all. In order to generalize from as much information as possible, we considered studies whether or not they tested for statistical significance of differences. We did this by simply rank-ordering the fields in terms of the information given in the study. This, of course, meant that sometimes the students in one field were ranked higher than the students in another even though the former were higher than the latter only by chance. This should not be disturbing, since the search is for trends across a number of studies, rather than for significant differences in any one study. If the two fields are indeed not different from each other, then by chance alone one field would probably be higher than the other in one study but lower in a second study. In short, chance differences are unlikely to "add up" to produce consistent results across studies. (This procedure, of course, presupposes a large enough number of studies to justify reliance on randomization.)

Since the different studies compare varying numbers of curricula, the rank of a field in one study could not be compared directly with the rank of the same (or similar) field in another study. Therefore ranks were "standardized" by dividing each rank in a study by the number of curricular groups in that study. By calculating the frequency (across studies) with which the standardized rank for

a field fell into the higher third of standardized ranks (.00-.33), the medium third (.34-.67), or the low third (.68-1.00), it was possible to determine something about the consistency of the findings across studies with respect to major-field differences for each of the several clusters of value, attitudinal, and personality variables.

In effect, then, a certain comparability among studies was achieved (1) by combining or clustering the independent and dependent variables of the studies at hand (major field, and student attributes, respectively), (2) by establishing a common "metric" for noting results (the ranks of major fields), and (3) by standardizing and comparing results (comparing "standardized" ranks).

The precision of the method used to compare the results of studies may vary. This variation is particularly well illustrated by Dubin and Taveggia (1968), who have collected a large number of studies, each of which reports one or more experimental comparisons of different teaching technologies (lecture only, discussion only, lecture with discussion, supervised independent study, and unsupervised independent study) with respect to the performance of students on course examinations. The integrators analyzed the results of these studies in four different ways; each successive method involved increasing precision but made use of a small number of comparisons in the original studies. The integrators first compared the results of studies by using a sign (direction) test across all comparisons in the original studies (for example, of 88 comparisons in 36 studies, 51 per cent favored instruction-by-lecture and 49% favored instruction-by-discussion with respect to student examination scores). In a smaller number of cases, where data on the distribution of scores (means, standard deviations, sample sizes) for the groups compared were available, they were able to use or calculate a standardized statistical measure of the differences in all possible mean examination scores; this procedure provides higher precision in determining the actual amount of differences between groups' scores in each comparison (a distribution of standardized differences is presented). Both of these procedures were then repeated for all studies where sufficient data were available, employing only independent comparisons. (In many of the studies a number of comparisons

were made in which the same groups were compared several times--on different exams at different periods during the semester; using specified criteria the integrators picked only one of these comparisons.) The sign test of these comparisons (now independent of one another) involved a smaller number of cases than in the first two comparative procedures. The final method of comparison--the distribution of standardized differences in measured group examination scores for independent comparisons--was the most precise but utilized the smallest number of cases.

Although there are techniques that can increase the comparability among studies in a given area, they are unlikely to produce total comparability. Studies will almost always vary in the settings investigated and the research instruments and procedures used. For instance, the studies of major-field differences, discussed above, are not uniform in many ways: they vary in the type of students sampled (college-class level of students), calendar year of the study, college studied, the research instrument used to measure a particular student characteristic. Moreover, although the use of standardized ranks solves the problem of varying number of curricula in various studies, it is not possible to adjust for the fact that investigators at different colleges have available for study different ranges and varieties of curricula. Moreover, even if two different colleges happened to have a similar range variety and range of curricula available for study, there is no assurance that investigators at the two schools would choose all or even the same curricula for study. In both cases, relative rankings (and thus standardized rankings) might be artificially inflated or deflated (see Feldman and Newcomb, 1969, Appendix G).

If despite the non-uniformity among studies, results are relatively consistent, the investigator has some assurance about the validity of his findings (unless it can be shown that the variability of research setting and method actually produced the consistencies artifactually.) More troublesome is the case when results are inconsistent across studies. It may be that inconsistencies are non-artifactual, indicating a need for articulation, specification, and revision of

theory. Here, the integration can be of most value by explicitly contrasting the theoretical orientations and research procedures of the studies. In other cases, the integrator may be able to resolve the inconsistencies by showing that in certain settings (say, multiversities) or for certain respondents (say, males), the results are rather consistent in a certain way, whereas results for some other setting (small colleges) or some other group of respondents (females) are also rather consistent, but in a different way. Or the investigator may want to weight the inconsistent studies differentially. It may be, for instance, that a large proportion of the studies do give consistent results, and the comparatively small number of studies at variance with these are primarily low-quality efforts whose results merit less than full consideration. It is always possible, however, that inconsistencies are really artifacts, due to the non-uniformities of the types described. If the integrator thinks this so, he is able to do little more than point it out, and urge further, more uniform, research in the area.

Closely related to the difficulties presented by non-uniformity are those due to missing data. The integrator must make do with what he is presented (not having had a hand in the original researches), which often is not enough. Lack of necessary data is not a unique problem of the integrator, but it is a distinctive one. Even those engaging in research-from-scratch ("original" or "primary" research) may find in the analysis stage of their research that there ^{are} ~~is~~ data they need but did not collect. Although the integrator is thus not alone in facing the situation of not having what he needs when he needs it, the problem does hit him with a vengeance. Especially bothersome are those cases in which lack of data prevent [^] him from introducing controls in order to specify or elaborate upon associations he has found. To take a brief example from the studies of major-field differences, it may be noted that most of these researches either do not present information about, or do not systematically take account of, the students' general background (socioeconomic status, gender, etc.). Therefore, the integrator cannot determine from these studies whether differences among major

fields with respect to students' values, attitudes and personality traits exist because of or in spite of differences in students' backgrounds. Do these differences reflect, say, merely differential distribution of the sexes in the several academic fields, or do they reflect the differential attraction of individuals differing in values and attitudes, and subsequent normative impact of the field on these students, for male and female students considered separately?

I want to end this section by noting the mundane--but nevertheless real problem--that arises when the integrator reaches the stage of reporting results of his analyses: what (and how much) should be presented where? The goal here is to strike a balance between the often conflicting requirements of completeness or thoroughness, on the one hand, and manageability and readability, on the other. In some cases the integrator will decide to describe and discuss a research report (or a part of it) in some detail in the text proper. He will do so for any of a variety of reasons: the study is a "classic" one; the research is of particularly high quality; the conceptual thinking underlying the research, in his opinion, is seminal; the study has a distinctive methodological approach or an especially good experimental design; the study contains useful information covered in very few (or no) other studies. Sometimes the integrator will make short summaries of the procedures and results of studies in a given area, and gather these summaries together in a "table," while offering generalizations in the text proper.⁴ These summaries, placed so as not to interrupt the flow of the text, not only save the reader the trouble of having to track down the studies summarized (assuming that brief summaries suffice), but also permit him to elaborate on, or check the validity of, the integrator's conclusions. In other cases, the integrator will offer generalizations of data in related studies--citing but not summarizing these studies. Finally, in still other instances, he will do no more than offer the reader a bibliographical list of materials on tangential topics not otherwise covered in his presentation.

FUNCTIONS AND VALUES OF INTEGRATIONS

Large-scale integrations have many values, not all of which are immediately apparent. Perhaps the most obvious use of an integration is as a compendium of research, providing the many services of a compendium: gathering together previously dispersed (and often not generally or easily available) information; condensing and abstracting existing research; providing bibliographic listings of research and theory. Furthermore, the mere existence of a compendium--if it is reasonably thorough--highlights the amount of research that has or has not been done. Whether or not the integrator comments on the matter, a review and integration makes evident which topics of interest have enjoyed little or no research and which bear a surfeit.

Integrations can be used to simulate a national sample. Almost any social scientific area benefits by having a broadly-based sample of respondents. Studies with such samples, being expensive, time-consuming, technically complex, and hard to execute, have been (and perhaps always will be) rare. Existing nationwide or broadly-based studies may be supplemented by reviews and integrations that piece together information from a number of "local" studies. (In the college effects area, this involves, for example, the piecing together of information from different studies done at different colleges throughout the United States; one problem, here, might be that the kinds of colleges studied produce a very biased sample.) Although not a substitute for more comprehensive studies, the juxtaposition of local studies does begin to provide some of the same advantages.

Another use of integration can be to simulate replication. The assertion that replication in social science is honored more in breach than in practice may be a cliché, but it is nevertheless true. The integrator can help close this breach. Although he is extremely unlikely to find studies that perfectly duplicate one another, he will usually find near-replicates. Thus he is able to determine whether any particular conclusion continues to hold under near-similar conditions.

An integration can be called upon to do many of the things that "primary" research does: testing hypotheses; lending (or not lending) support to extant

proposition; offering new information and generalizations; generating theoretical issues to be explored; and suggesting future research. I again turn to the integration in which Theodore M. Newcomb and I were involved for illustrations. The first example primarily describes an attempt to corroborate an extant proposition; the second focusses on an attempt to create generalizations. Both examples involve hypotheses testing and suggestions for future research.

Based on quite limited evidence, Jacob (1957) concluded that "there is more homogeneity and greater consistency of values among students at the end of their four years than when they begin" (p.4). Freshman-senior differences in standard deviations can be used as one means of determining whether colleges typically have the impact of "homogenization" rather than "heterogenization." Across colleges, in almost every change-area, decreases in homogeneity are about as likely as increases. Put otherwise, increasing homogeneity of outlook among students is a variable outcome of the college experience rather than a constant one. This finding, then, leads to an inquiry about the conditions under which increasing homogeneity is more likely than decreasing homogeneity (and vice versa). Many hypotheses can be offered--although I shall not pursue them here--all calling for future research.

Propositions can be generated, and corroborating evidence sought, by performing "new" statistical operations on data given in studies. Such re-analysis is, in a sense, a type of "secondary analysis," although it does not involve, as is more commonly the case, acquiring some other researcher's data decks and then doing one's own analyses. In searching out the effects of major fields on students, we did a rather simple re-analysis of data from a number of studies, taking our initial clue from a study by Huntley (1965). His longitudinal investigation traced changes of students' scores on the Allport-Vernon-Lindzey Study of Values, contrasting their scores at entrance with those obtained just prior to graduation. Separate data are presented and analyzed for each of nine curricular groups. Almost in passing, he notes that "the initial differences among the groups tend to be accentuated or sharpened over the four years" (p.381),

showing this by comparing the appropriate F statistics: for each of the six Spranger values, the size of the F statistic is larger for the groups at graduation than for the groups at entrance.

We would have liked to compare the F statistics for "entrance" and "exit" major-field groups in other studies presenting data on change and stability of students in various academic majors. Since the studies with appropriate entrance-exit data did not present the necessary F statistics, an alternative strategy had to be employed to see whether Huntley's finding was generalizable: a rank-order correlation between rank of initial major-field average score and rank of the average gain in score was calculated. It was found that, in general, accentuation of initial major-field differences (as indicated by a positive rank-order correlation between ranking of initial average scores of curricula and ranking of average gains) is much more likely than is the minimization of such differences (indicated by a negative rank-order correlation). The accentuation phenomena suggests that the processes of selection into a major field are interdependent with the processes involved in the impacts of these fields on their students. Additional research is needed to establish the validity of this suggestion as well as to determine such things as (1) the conditions under which accentuation is most likely to occur and (2) the exact contribution of environmental pressures of the major field, ~~the~~ the importance ^{of} personality dynamics, and the mechanisms of interplay between these two sets of influences.

Any study, or set of studies, benefits when put into the perspective provided by the results of other relevant studies. It is this benefit of embedding a particular study in the empirical and theoretical framework provided by other studies that integrations, in their search for generalizations and specifications, are in such a good position to supply. I should like to expand on this ^{particular} function of integrations by viewing it in the context of the degree of criticalness in which integrations engage.

The integrator must navigate between the Scylla and Charybdis of "hyper-criticalness" and "hypocriticalness." No study in the behavior ^{al} sciences is

totally perfect, definitive, and complete. No research instrument is totally faultless. It would be possible to set one's critical standards so high that no research or research instrument would pass muster. One would reject the findings of all studies in a particular domain, coming to the patently absurd conclusion that nothing is really known about the area. Total rejection, or anything approaching it (hypercriticalness), may unjustly undermine the research efforts and knowledge of a field. On the other hand, to view and present all studies and research instruments as perfect unto themselves, and thus of equal worth, contributes little to the field, at best, and might well generate seriously incorrect information, at worse. Justice must be tempered with mercy, but mercy must also be tempered with justice.

The integrator has plenty of occasions to use his critical facilities. Part of his decision about what to emphasize where in his written report depends upon the perceived differential quality of the materials under consideration. As previously suggested, these differentials also may help the integrator resolve certain inconsistencies in the results of studies in the same area. Moreover, the integrator is obligated to point out deficiencies as well as strengths of research instruments and investigations, even in those cases--perhaps, the more felicitous phrase is 'especially in those cases'--of high-quality researches that have advanced the state of the arts in the field. It is also important for him to point out what he believes to be incorrect interpretations or over-interpretations of results, perhaps offering more satisfactory interpretations. Still, the stance of criticalness can be pushed too far. I have heard it argued that in any social scientific field at least ninety per cent (and probably more) of the studies are of such low quality as to be virtually worthless.⁵ The results of these studies can be ignored without loss; to try to integrate them is at best fruitless, and at worse misleading. Another argument uses similar premises to arrive at a different suggestion. Why not, it is asked, rate all studies in a field as to their quality? The lowest ninety per cent could be listed and then eliminated from further consideration. The integration would consider only the remaining ten per cent, and best, studies.

One could indeed rate all studies in a field, but it would not be an easy matter; and I am not sure how valid the results would be nor whether the effort

would be worth the candle. In the first place, there are a number of dimensions along which quality may vary. One study might be especially exciting theoretically but weak in statistical analysis of the collected data. A second study may be especially sophisticated in data analysis but contain almost no theoretical orientation. A third study may have used particularly good techniques to generate data but is weaker than the first two in both theory and data analysis. Some way of arriving at a meaningful global rating (across these and other dimensions) would have to be constructed. This may or may not be possible. Even if it could be done, I doubt that the general intercoder reliability would be very high. This implies, of course, that at least two-- and preferably more ⁻⁻persons independently read and rate all or some proportion of the studies, which is not always feasible.⁶

It is possible, that even though there is low agreement about the exact placement of studies, judges would show fairly high agreement about which studies fall into the top decile. Supposing this to be the case I would still argue that to consider only these top studies reduces perspective and in some instances can be misleading. Not everyone agrees with the low estimate of the quality of work done in the behavioral science domain. It seems to me that a substantial proportion of studies are of high enough quality to be of direct use. I would even go a step further, and maintain that the results of lesser-quality research, if used correctly, can be of service. The most useful strategy is not to throw out nearly the lion's share of available information, nor to accept everything uncritically, but to use what is available judiciously.

Imagine that a researcher at an elite college has found--in a high-quality study--that, on the average, students as seniors are less authoritarian, dogmatic, and prejudiced than they were as freshmen. It is helpful to the investigator (as well as to the integrator) to know that virtually all other relevant studies--of high-quality or otherwise, with a sophisticated research design or otherwise, at a wide variety of elite and non-elite schools--show average decreases in these traits during the college years. Knowing this, the research can with some con-

confidence rule out the "eliteness" of the school as a cause of decreases per se (although it may be that the magnitude of decrease may still be affected).

It is possible that the results of a well-known, relatively highly-regarded study may be unintentionally misleading. I think that one aspect of the longitudinal study done at Vassar in the early 1950s (see Freedman, 1961, 1967; Sanford, 1956, 1962, 1967) provides a case in point. The "Vassar Study" is generally recognized as one of the larger and better quality researches in the area of college impacts. One of the conclusions of this study is that the impact of college is greatest during the earlier part of the college experience. Basic or major changes in college, it is maintained, occur early in the college experience because of the special sensitivity of freshmen and sophomores to the influences they encounter. Juniors and seniors are considered to be in a different developmental phase, one where change is leveling off and where little more is happening. These views are now quite widely held, as appropriate to students at most (perhaps all) colleges; their prevalence is due in large part, I would say, to the persuasiveness of the Vassar researchers and the quality of their efforts. Still there are grounds for not expecting to find, as an invariable occurrence at most colleges for most colleges for most individual characteristics, that college effects are greatest during the first year or two. (see Feldman and Newcomb, 1969, Ch. 4).

To provide more information in this area, contiguous college-class levels can be compared with respect to average difference-scores on a variety of value, attitudinal, and personality indicators. This is not the only nor necessarily the best way of getting at timing of impact; but it is one way, and it permits the use of information from a large number of studies. Support would be lent to the early-impact-and-change proposition if freshman-sophomore differences (in means) for most studies are larger than either sophomore-junior or junior-senior differences. It turns out that with one major exception, this is not the case. The major exception is for authoritarianism (and related attributes) where freshman-sophomore differences (decreases) are larger than sophomore-junior

and junior-senior differences in the majority of investigations. Most probably it is more than coincidental that many of the scales used in the Vassar Study were either authoritarianism scales (or near relatives) or contained a number of authoritarianism-like items. Thus it might be argued that because of the nature of many of the scales used in the Vassar Study, and perhaps because of the particular nature of the sample (Vassar girls), the results with **respect to timing of impact, although correct for Vassar at the time in particular** content areas, have been too readily generalized to other schools and to other areas. Systematic consideration of other relevant studies (of high and lesser quality) help put into perspective the results of this study.

Many of the values of integrations listed so far can be subsumed under the very general value of assisting in the cumulation of (social) science. If one believes in the cumulative nature of science, then periodic stock-taking becomes essential for any particular arena of scientific endeavor. The cumulation of knowledge in an area may, of course, occur more or less haphazardly--but this does not, and should not, preclude more systematic attempts by laborers in a field to determine where they have arrived (and, consequently, where they might go). It is as this kind of systematic attempt that makes large-scale reviews and integrations important.

It may be argued that social science does not so much cumulate as evolve. If a "natural" evolution is allowed, only a small proportion of the studies--the best or the "fittest"--would survive. Those studies that do not "die," and only those, deserve to remain--so the argument runs. The implication here (to put the matter in its most extreme) is that integrations, in their misplaced effort to aid in the cumulation of social science, actually hinder its evolution. But I see no reason to expect that those studies that survive are necessarily the "fittest," since there is not complete agreement among researchers about the quality of any given study, and since, in any case, "scientific" quality is probably not the sole determinant of survival. (The suggestion that the integrator eliminate from consideration ninety per cent of studies in a field is

a variant of the "survival of the fittest argument": the integrator is asked to pick the "fittest," rather than relying on more "natural" processes.) Even if those that survived were the best, the information and results of less fit studies may still serve an important use. As elaborated above, taking into consideration "all" studies in an area can provide needed perspectives and correct misleading impressions.

The cumulation versus evolution distinction should not be made too rigidly, however, as though only one or the other of these completely characterize^A a science. If Thomas S. Kuhn is correct in "The Structure of Scientific Revolutions," a science "progresses" by cumulating within its established paradigm(s) and by evolving as a consequence of conflicts among paradigms. In his conceptualization, there is both "normal" science and research (within a paradigm) and "revolutionary" science and "extraordinary" research (on those occasions when a new paradigm arises to challenge and replace another). Within this framework, the integrator's efforts fall primarily within "normal" research:

Normal science, the puzzle-solving activity we have just examined, is a highly cumulative enterprise, eminently successful in its aim, the steady extension of the scope and precision of scientific knowledge (p.52).

...scientific revolutions are here taken to be those non-cumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one (p. 91).

An integration does not ordinarily concentrate on constructing a new paradigm to challenge existing ones.⁷ In this sense, an integration occupies a basically "conservative" position, as part of "normal" science. This term describes, not condemns, integrations, for, as Kuhn take pains to show, both normal (conservative) and revolutionary (radical) scientific efforts have functions and dysfunctions. Moreover, it should be realized that integrations can help to precipitate "scientific revolutions" by pointing out empirical anomalies, by suggesting counterinstances to established generalizations and exposing putative "truths" as "myths," by highlighting divergent articulations within a given research tradition, and by systematizing and clarifying the conflicts between schools of thought.

A good integration, at the same time that it is showing how much is known in an area, is also showing how little is known. It sums up, but does not end. In this sense, it is only a beginning.

FOOTNOTES

¹This article deals with large-scale reviews and integrations, in which the investigator defines his research to be the integration of others' efforts rather than the collecting and analysis of "new" data. Integration, as a research method, has both similarities and dissimilarities to such related processes as more limited review of the literature as a prelude to "original" research, secondary analysis of data already collected and initially analyzed, the use of data banks, and computerized retrieval and processing of information (Lindvahl, 1959; Bisco, 1966; Selvin, 1968). It is beyond the scope of this paper, however, to compare and contrast large-scale integration with these related research techniques.

²Particularly useful to researchers interested in educational processes are such references as the following: "Education Index"; "College Student Personnel Abstracts"; "Sociological Abstracts"; "Psychological Abstracts"; "Inventory of Current Research in Higher Education" (Heckman and Martin, 1968); "Current Contents"; "Encyclopedia of Educational Research" (Harris, 1960); "Handbook of Research on Teaching" (Gage, 1963); "Sources in Educational Research" (Manheim, Dardarian, and Satterthwaite, 1969); the Educational Resources Information Center's "Research in Education," "Current Index to Journals in Education," "Selected Documents in Higher Education," and the bulletins, bibliographies, and research reviews of the several clearinghouses in the ERIC system; and the services of the Science Information Exchange in Washington, D.C.

³See Hoffman (in press) as an example of how directions of results--in this case the association of types of parental childrearing practices and moral development indices--from several studies can be tallied and a theoretical interpretation proffered to fit the general pattern of results.

⁴Such "tables" or "charts" have been used to good effect in Calhoon and Reddy (1968), Hyman (1959), and Murphy, Murphy, and Newcomb (1937); also see Feldman and Newcomb (1969).

⁵It may be noted that the domain of educational research in particular has been found wanting. See Levit (1968) for references documenting that "almost all major aspects of educational research has been widely and severly criticized" (p. 145).

⁶For a good analysis of the problems involved in a critical examination and rating of studies--including the multidimensionality and the "mutability" of methodological criteria--see Hirschi and Selvin (1967).

⁷A paradigmatic challenge may contain a good overall review of certain works in a field, and may contain many bibliographical citations, but it is not conceived as a full-fledged integration of a field. Consider, as a case in point, Kohlberg's (1969) elaboration of the "cognitive-developmental" approach to socilization as a challenge to the psychoanalytic and social-learning approaches.

REFERENCES

- Bisco, Ralph L. Social Science Data Archives: A Review of Developments. American Political Science Review 60: 93-109; 1966.
- Calhoon, Richard A. and Reddy, A. C. The Frantic Search for Predictors of Success: 50 Years of Confusion and Contradiction. Journal of College Placement 28: 54-66; 1968.
- Dubin, Robert and Taveggia, Thomas C. The Teacher-Learning Paradox: A Comparative Analysis of College Teaching Methods. Eugene, Oregon: Center for the Advanced Study of Educational Administration, University of Oregon, 1968.
- Feldman, Kenneth A. and Newcomb, Theodore M. The Impact of College on Students. Vol. 1. An Analysis of Four Decades of Research. Vol. 2. Summary Tables. San Francisco: Jossey-Bass, 1969.
- Freedman, Mervin B. Measurement and Evaluation of Change in College Women. U.S. Dept. of Health, Education, and Welfare Cooperative Research Project No. 736. Poughkeepsie, N.Y.: Mellon Foundation, Vassar College, 1961.
- Freedman, Mervin B. The College Experience. San Francisco: Jossey-Bass, 1967.
- Gage, Nathaniel L. (editor). Handbook of Research on Teaching. Chicago: Rand McNally and Co., 1963.
- Harris, Chester W. (editor). Encyclopedia of Educational Research. Third edition. New York: McGraw-Hill Book Co., 1959.
- Heckman, Dale M. and Martin, Warren B. Inventory of Current Research on Higher Education 1968. New York: McGraw-Hill Book Co., 1959.
- Hirschi, Travis and Selvin, Hanan C. Delinquency Research: An Appraisal of Analytic Methods. New York: Free Press, 1967.
- Hoffman, Martin L. "Socialization Practices and the Development of Moral Character." Character Development in the Child. (Edited by Martin L. Hoffman.) Chicago: Aldine, in press.
- Huntley, C. William. Changes in Study of Values Scores during the Four Years of College. Genetic Psychology Monographs 71: 349-383, 1965.
- Hyman, Herbert H. Political Socialization: A Study in the Psychology of Political Behavior. New York: Free Press, 1959.
- Kohlberg, Lawrence. "Stage and Sequence: The Cognitive-Developmental Approach to Socialization." Handbook of Socialization Theory and Research. (Edited by David A. Goslin.) Chicago: Rand McNally and Co., 1969. Chapter 6, pp. 347-480.
- Kuhn, Thomas S. The Structure of Scientific Revolutions. International Encyclopedia of Unified Science, Foundations of the Unity of Science, Vol. 2, No. 2. Chicago: University of Chicago Press, 1962.
- Levit, Martin. Scientific Method, Social System, and Educational Research. Studies in Philosophy and Education 6: 145-167; 1968.

Lindvall, C. M. The Review of Related Research. Phi Delta Kappan 40: 179-180; 1959.

Manheim, Theodore; Dardarian, Gloria L.; and Satterthwaite, Diane A. Sources in Educational Research: A Selected and Annotated Bibliography. Part I. General. Detroit: Wayne State University Press, 1969.

Sanford, Nevitt (issue editor). Personality Development during the College Years. Journal of Social Issues 12: 1-71; October 1956.

Sanford, Nevitt (editor). The American College: A Psychological and Social Interpretation of the Higher Learning. New York: Wiley, 1962.

Sanford, Nevitt. Where Colleges Fail: A Study of the Student as a Person. San Francisco: Jossey-Bass, 1967.

Selvin, Hanan C. "The Computer Analysis of Observational Data." Calcul et Formalisation dans les Sciences de L'homme. Paris: Centre National de la Recherche Scientifique, 1968. Pp. 228-238.